Chemometrics and Intelligent Laboratory Systems 133 (2014) 25–32

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemolab

# Multi-way PLS for discrimination: Compact form equivalent to the tri-linear PLS2 procedure and its monotony convergence

Samia Samar Ouertani [a], Gérard Mazerolles [c], Julien Boccard [b], Serge Rudaz [b], Mohamed Hanafi [a,*]

[a] ONIRIS, Unité de Sensométrie et de Chimiométrie, site de la Géraudière, Nantes, France
[b] School of Pharmaceutical Sciences, University of Genève, University of Lausanne, Geneva, Suisse
[c] INRA, UMR 1083, Montpellier, France

ARTICLE INFO

*Article history:*
Received 23 July 2013
Received in revised form 29 January 2014
Accepted 31 January 2014
Available online 7 February 2014

*Keywords:*
Multi-way PLS regression
Discriminant analysis
Tri-linear PLS2
Three-way data

ABSTRACT

This paper focuses on multi-way PLS for discrimination. Emphasis was placed on the computation of parameters (so-called scores and loadings) using an iterative procedure called tri-linear PLS2. In the context of discrimination, this algorithm is applied to a dummy matrix representing group membership, and according to its specific formalism, the tri-linear PLS2 procedure offers the possibility of simplification. The purpose of this paper is to introduce a compact formulation of the tri-linear PLS2 procedure adapted for a discrimination setting. A property of this variant that allows formally establishing its convergence and, by extension, the convergence of the tri-linear PLS2 will be demonstrated. The potential of this compact form will be illustrated with simulated examples.



## 1. Introduction

Chemometricians frequently address discrimination or classification issues. Usually, the dataset to be analyzed can be viewed as a matrix where $I$ samples (or rows) are divided into G classes or groups related to an (I, 1) categorical variable **y**. The $J$ columns of **X** describe measurements or quantitative variables. For convenience, the rows of **X** can be permuted such that the $I_1$ first rows correspond to the first group and the following $I_2$ rows correspond to the second group, etc.

Partial Least Squares Discriminant Analysis (PLS-DA) is a very popular method to discriminate or classify different groups of samples. PLS-DA consists of a classical PLS regression model [1] in which the multivariate response data are replaced by a dummy matrix **Y** that describes the categories. The dummy matrix defines the class membership of the statistical units; i.e. the matrix **Y** is of dimension I × G, where G is the total number of groups and $y_{ig}$ is equal to 1 if the ith observation belongs to gth group and is equal to 0 otherwise.

PLS-DA was properly formalized by Barker and Rayens [2], who provided a formal statistical background to describe the PLS regression properties in the particular case of discrimination. Subsequently, extensions and variants of PLS-DA were then proposed [3–6] and have been widely used in various application areas, such as food science, image analysis, and process monitoring [7–13].

Currently, data generated by modern analytical devices such as separation technique coupled with mass spectrometry detection and advanced spectroscopic approaches are often very large. In some cases, these data can be summarized meaningfully in a multi-dimensional data structure called multi-way data or tensor. For example, monitoring biological or chemical phenomena over time is especially important in the life sciences. The data generated (or observed compounds) can classically be arranged in a three-way table (or third-order tensor) of size (I, J, K) consisting of I samples (or individuals) described by J compounds (or detected analytes) and measured at K different points in time [14] (see Fig. 1).

When different situations or modalities have to be compared, the most common procedure for classifying samples from the original structure is to unfold (or matricize) the three-way data into two-way arrays and apply traditional multivariate tools for classification, typically PLS-DA. However, altering the organization of the data in this way introduces the risk of losing information. Recently, several works have used an alternative approach that is applied directly on a three-way table to avoid the unfolding process [15–17]. This modeling strategy is based on multi-way Partial Least Squares regression (N-PLS), which was introduced by Bro [18]. For discrimination, a dummy matrix of groups is used as a response variable, as previously described for the conventional two-way case. Similar or better results were obtained with N-PLS in comparison with the unfolding strategy. More importantly a significant gain in terms of interpretation was reported because there were fewer parameters to evaluate [14].

* Corresponding author.
*E-mail addresses:* ouertani.samiasamar@yahoo.fr (S.S. Ouertani), gerard.mazerolles@supagro.inra.fr (G. Mazerolles), Julien.Boccard@unige.ch (J. Boccard), Serge.Rudaz@unige.ch (S. Rudaz), mohamed.hanafi@oniris-nantes.fr (M. Hanafi).

0169-7439/$ – see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.chemolab.2014.01.015

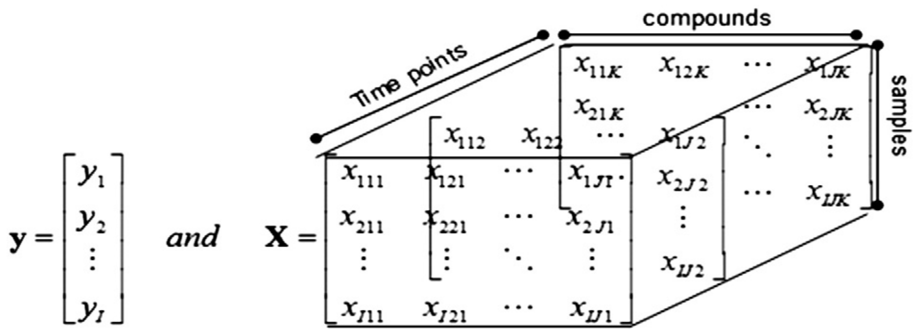


**Fig. 1.** Classical three way data **X** generated in time series.

The main issue motivating this study is the computation of N-PLS parameters (scores and loadings) in the specific case of discrimination (N-PLS-DA), based on an iterative procedure called tri-linear PLS2 [18]. In this context, the procedure is applied to a dummy matrix **Y**, and the tri-linear PLS2 procedure offers the possibility of simplifications according to its particular form.

This paper introduces a compact form of the tri-linear PLS2 procedure that is adapted to the discrimination Scheme. A monotony property of the compact formulation is demonstrated and shows its convergence and therefore, the convergence of the tri-linear PLS2. These contributions are illustrated using simulated examples.

This paper is organized as follows: Section 2 summarizes notations and basic definitions from tri-linear algebra; Section 3 provides a brief presentation of Bro's multi-way PLS regression; and Section 4 focuses on the discrimination case and introduces the compact form of the tri-linear PLS2 procedure. The properties of this compact procedure in terms of monotony and convergence are also established.

## 2. Notations

Various notations have been used to describe the multi-way PLS regression method. Bro's presentation is based on matrix notation, while Smilde [19] uses Kronecker product notation [20]. This paper follows a description of the multi-way PLS regression method based on n-mode product notation [21,22]. This choice appears to be more rigorous and uses fewer indexes compared to the others methods of notation.

This section summarizes the basic definitions and notations required in subsequent developments. Matrices and third-order tensors are in uppercase bold and vectors are in lowercase bold.

### 2.1. Basic definitions

- Let **a**, **b** and **c** be the three real vectors of size (I × 1), (J × 1) and (K × 1) respectively, a third-order tensor **X** is rank one if it can be written as the outer product of three vectors, i.e., $\mathbf{X} = \mathbf{a} \bigcirc \mathbf{b} \bigcirc \mathbf{c}$. The symbol "○" represents the vector outer product. This means that each element of the tensor is the product of the corresponding vector elements: $x_{ijk} = a_i b_j c_k$.
- Let **X** and $\widetilde{\mathbf{X}}$ be the two tensors of size (I, J, K). The scalar product of two third-order tensors is denoted by $\left\langle \mathbf{X}, \widetilde{\mathbf{X}} \right\rangle$ and computed as a sum of element-wise products over all indices; that is, $\left\langle \mathbf{X}, \widetilde{\mathbf{X}} \right\rangle = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} x_{ijk} \widetilde{x}_{ijk}$. The scalar product allows the norm of a third-order tensor **X** to be defined as:

$$\|\mathbf{X}\| = \sqrt{\langle \mathbf{X}, \mathbf{X} \rangle} = \sqrt{\sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} x_{ijk}^2}.$$

### 2.2. n-mode products

To multiply a third-order tensor by a vector or a matrix, it is necessary to specify the corresponding tensor mode (or way).

1- The 1-mode product $\mathbf{X} \overline{\times}_1 \mathbf{u}$ of a third-order tensor **X** of size (I, J, K) by an I-vector **u** is a matrix **M** of size (J, K) with elements:

$$m_{jk} = \sum_{i=1}^{I} x_{ijk} u_i (1 \le j \le J) \text{ and } (1 \le k \le K).$$

In the same manner, the 2-mode product $\mathbf{X} \overline{\times}_2 \mathbf{u}$ of a third-order tensor **X** of size (I, J, K) by a J-vector **u** is a matrix **M** of size (I, K) with elements $m_{ik} = \sum_{j=1} x_{ijk} u_j$. Similarly, the 3-mode product $\mathbf{X} \overline{\times}_3 \mathbf{u}$ of a third-order tensor **X** of size (I, J, K) by a K-vector **u** is a matrix **M** of size (I, J) with elements $m_{ij} = \sum_{k=1}^{K} x_{ijk} u_k$ (see Fig. 2 for a graphical representation of the n-mode product).

2- The 1-mode product $\mathbf{X} \times_1 \mathbf{U}$ of a third-order tensor **X** of size (I, J, K) by a matrix **U** of size (L, I) is a third-order tensor $\hat{\mathbf{X}}_1$ of size (L, J, K) with elements:

$$\hat{x}_{ljk} = \sum_{i=1}^{I} x_{ijk} u_{li} \text{ with } (1 \le l \le L), (1 \le j \le J) \text{ and } (1 \le k \le K).$$

In the same manner, the 2-mode product $\mathbf{X} \times_2 \mathbf{U}$ of a third-order tensor **X** of size (I, J, K) with a matrix **U** of size (L, J) is a third-order tensor $\hat{\mathbf{X}}_2$ of size (I, L, K) with elements $\hat{x}_{ilk} = \sum_{j=1}^{J} x_{ijk} u_{lj}$. Similarly, the 3-mode product $\mathbf{X} \times_3 \mathbf{U}$ of a third-order tensor **X** of size (I, J, K) with a matrix **U** of size (L, K) is a third-order tensor $\hat{\mathbf{X}}_3$ of size (I, J, L) with elements $\hat{x}_{ijl} = \sum_{k=1}^{K} x_{ijk} u_{lk}$ (see Fig. 3 for a graphical representation of the n-mode product).

## 3. Brief presentation of Bro's N-PLS regression

The following definition is necessary to explore the systematic variation patterns in a three-way data set **X** of size (I, J, K), which are likely to predict the systematic variation patterns in **Y**, a response data matrix of size (I, Q) (**Y** is not necessarily a dummy matrix). The multi-way PLS regression model for the mean-centered data **X** and **Y** is defined as:

$$\mathbf{X} = \sum_{h=1}^{H} \mathbf{t}_{\mathbf{X},h} \circ \mathbf{a}_{\mathbf{X},h} \circ \mathbf{b}_{\mathbf{X},h} + \mathbf{R}_{\mathbf{X}}^{(H)} \quad (1)$$

$$\mathbf{Y} = \mathbf{T}_{\mathbf{X}}^{(H)} \mathbf{B}^{(H)} + \mathbf{R}_{\mathbf{Y}}^{(H)} \quad (2)$$

In Eq. (1) $\mathbf{t}_{\mathbf{X},h}$ denotes the so-called X-scores; $\mathbf{a}_{\mathbf{X},h}$ and $\mathbf{b}_{\mathbf{X},h}$ indicate the **X**-loadings associated with the second and third mode, and $\mathbf{R}_{\mathbf{X}}^{(H)}$ denotes the residual part.